ATSC Standard:
Audio Watermark Emission

Doc. A/334:2022-03
31 March 2022

Advanced Television Systems Committee
1300 I Street, N.W., Suite 400E
Washington, D.C. 20005
202-872-9160

The Advanced Television Systems Committee, Inc. is an international, non-profit organization developing voluntary standards and recommended practices for broadcast television and multimedia data distribution. ATSC member organizations represent the broadcast, professional equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. ATSC also develops implementation strategies and supports educational activities on ATSC standards. ATSC was formed in 1983 by the member organizations of the Joint Committee on Inter-society Coordination (JCIC): the Consumer Technology Association (CTA), the Institute of Electrical and Electronics Engineers (IEEE), the National Association of Broadcasters (NAB), the Internet & Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). For more information visit www.atsc.org.

Note: The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Implementers with feedback, comments, or potential bug reports relating to this document may contact ATSC at https://www.atsc.org/feedback/.

Revision History

Version	Date
Candidate Standard approved	27 November 2015
Standard approved	19 September 2016
Reference [2] updated to point to published version of A/342 Part 1:2017	24 February 2017
Reference [3] updated to point to published version of A/336:2017	27 February 2017
A/334:2022-03 published (references to ATSC documents updated)	31 March 2022

Table of Contents

1. SCOPE 1
1.1 Introduction and Background 1
1.2 Organization 1
2. REFERENCES 1
2.1 Normative References 1
2.2 Informative References 1
3. DEFINITION OF TERMS 1
3.1 Compliance Notation 1
3.2 Treatment of Syntactic Structures 2
3.2.1 Reserved Fields 2
3.3 Acronyms and Abbreviation 2
3.4 Terms 2
4. SYSTEM OVERVIEW 3
4.1 System Architecture 3
5. SPECIFICATION 4
5.1 Physical Layer 4
5.1.1 Marked Audio Signal 4
5.1.2 Marked Audio Presentation 6
5.2 Data Link Layer 6
5.2.1 Cell 6

Index of Figures and Tables

Figure 4.1 Generic audio watermark system architecture..... 3

Table 5.1 Mapping of Autocorrelation Difference Value to Symbol and Bit Values 5
Table 5.2 Parameter Specifications 5
Table 5.3 Syntax cell() Structure 6
Table 5.4 Header Sequence 6

ATSC Standard: Audio Watermark Emission

1. SCOPE

1.1 Introduction and Background

This document specifies the VP1 audio watermark for use with systems conforming to the ATSC 3.0 family of specifications. This document specifies the format in which the audio watermark resides in a PCM audio signal. Emission by a broadcaster of the audio watermark is optional.

A specification for the use of VP1 audio watermarks in emissions with application to the recovery of ATSC 3.0 service signaling by redistribution receivers is provided in A/336 [3].

1.2 Organization

This document is organized as follows:

- Section 1 – Outlines the scope of this document and provides a general introduction.
- Section 2 – Lists references and applicable documents.
- Section 3 – Provides a definition of terms, acronyms, and abbreviations for this document.
- Section 4 – System overview
- Section 5 – System specifications

2. REFERENCES

All referenced documents are subject to revision. Users of this Standard are cautioned that newer editions might or might not be compatible.

2.1 Normative References

The following documents, in whole or in part, as referenced in this document, contain specific provisions that are to be followed strictly in order to implement a provision of this Standard.

[1] IEEE: "Use of the International Systems of Units (SI): The Modern Metric System," Doc. SI 10, Institute of Electrical and Electronics Engineers, New York, NY.

[2] ATSC: "ATSC Standard: Audio Common Elements," Doc. A/342-1:2022-03, Advanced Television Systems Committee, Washington, DC, 31 March 2022.

2.2 Informative References

The following documents contain information that may be helpful in applying this Standard.

[3] ATSC: "ATSC Standard: Content Recovery in Redistribution Scenarios," Doc. A/336:2022-03, Advanced Television Systems Committee, Washington, DC, 31 March 2022.

3. DEFINITION OF TERMS

With respect to definition of terms, abbreviations, and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards [1] shall be used. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question will be described in Section 3.3 of this document.

3.1 Compliance Notation

This section defines compliance terms for use by this document:

shall – This word indicates specific provisions that are to be followed strictly (no deviation is permitted).

shall not – This phrase indicates specific provisions that are absolutely prohibited.

should – This word indicates that a certain course of action is preferred but not necessarily required.

should not – This phrase means a certain possibility or course of action is undesirable but not prohibited.

3.2 Treatment of Syntactic Structures

This document contains symbolic references to syntactic structures used in the audio, video, and transport coding subsystems. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

3.2.1 Reserved Fields

One or more reserved bits, symbols, fields, or ranges of values may be present in this document. These are used primarily to enable adding new values to a syntax structure without altering its syntax or causing a problem with backwards compatibility, but they also can be used for other reasons.

The ATSC default value for reserved bits is '1.' There is no default value for other reserved fields. Use of reserved fields except as defined in ATSC Standards or by an industry standards setting body is not permitted. See individual field semantics for mandatory settings and any additional use constraints. As currently-reserved fields may be assigned values and meanings in future versions of this Standard, receiving devices built to this version are expected to ignore all values appearing in currently-reserved fields to avoid possible future failure to function as intended.

3.3 Acronyms and Abbreviation

The following acronyms and abbreviations are used within this document.

ATSC – Advanced Television Systems Committee

LSB – Least significant bit.

MSB – Most significant bit.

bslbf – Bit string, left bit first.

3.4 Terms

The following terms are used within this document.

audio presentation – Has the meaning given in the ATSC A/342 Audio, Part 1: Common Elements [2]. Also referred to as a Preselection(DASH-IF), a Presentation(AC-4), or a Preset(MPEG-H).

audio signal – Has the meaning given in the ATSC A/342 Audio, Part 1: Common Elements [2].

audio watermark – Data which is embedded in audio essence in such a way that it can be extracted (i.e., read) by an appropriately designed extractor.

embed – The process whereby an audio signal is modified to include an audio watermark.

embedder – A tool or process that is able to embed an audio watermark in an audio signal.

extractor – A tool or process that is able to extract audio watermark packets from an audio signal.

cell – A complete transmission of an independently recoverable packet of data in an audio watermark.

header – A sequence of bits that is present at the start of a cell.

marked audio – Audio that has an audio watermark embedded in it.

PCM – Linear Pulse Code Modulation. In this document this is understood to be the uncompressed format for audio signals.

symbol – The representation of a bit of binary information in the audio watermark.

VP1 – The audio watermarking technology standardized in this specification

4. SYSTEM OVERVIEW

4.1 System Architecture

Figure 4.1, below, is a block diagram of the basic architecture for use of VP1 audio watermark technology. At the left of the figure, prior to distribution of content to receivers, an audio presentation is input to a VP1 audio watermark embedder, along with a sequence of data packets intended for delivery to a receiver via the VP1 audio watermark. The output of the embedder is a marked audio presentation, which is included in distribution content. Distribution content is delivered to receivers which incorporate a VP1 audio watermark extractor. When a marked audio presentation is present in the received audio, the extractor recovers the embedded data packets from the VP1 audio watermark.

The VP1 audio watermark is co-resident with audio energy in a region of the audio frequency spectrum containing perceptually important components and is therefore retained through audio distribution paths. It employs a modulation and encoding scheme which is resilient to most types of audio processing, including modern techniques for lossy signal compression. It also includes error detection and correction capabilities. Together, these features permit VP1 to provide a reliable auxiliary channel for delivery of a data payload to accompany audio and audiovisual content through environments employing heterogeneous audiovisual formats, protocols, and interfaces.

Audio watermark embedding is expected to be performed in broadcast environments on PCM audio prior to emission encoding or at an upstream transcode point. For audio delivered to receivers in an encoded format, audio decoding to PCM audio is expected to be performed prior to audio watermark extraction.

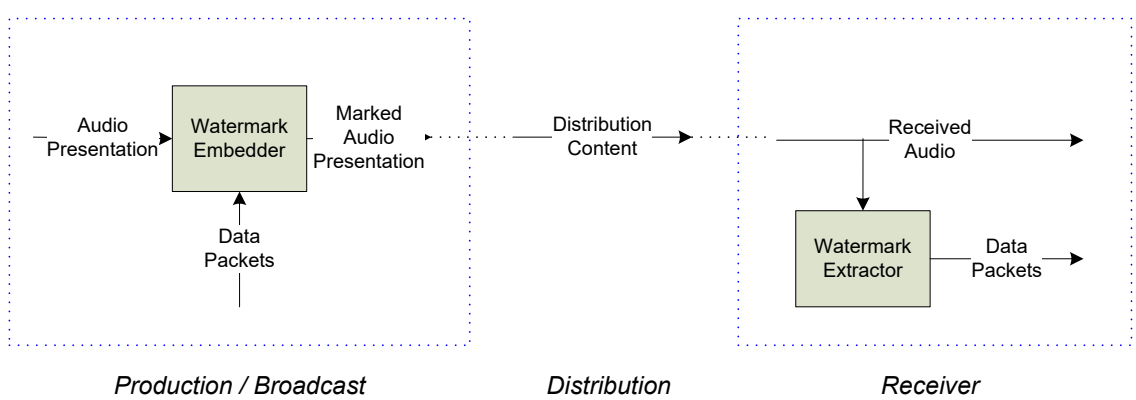

Figure 4.1 Generic audio watermark system architecture.

The VP1 audio watermark is specified independently of the format in which the audio presentation is produced; e.g. channel-based, object-based, ambisonic, etc. For any format, the audio presentation is understood to be comprised of one or more monophonic audio signals that are intended for synchronized, simultaneous rendering. The audio watermark is specified to be present in all synchronous audio signals, with the same symbol embedded at the same rendering time of each audio signal. With this approach, embedded audio watermark data can be extracted from any individual audio signal from the audio presentation, from any subset of audio signals from the audio presentation, or from any linear combination of audio signals from the audio presentation. This approach ensures that the audio watermark can be recovered even when multiple audio signals are mixed together during decoding, rendering, and/or transcoding of the marked audio presentation prior to audio watermark extraction.

The VP1 audio watermark is specified in Section 5 as a sequence of contiguous symbols that comprise a contiguous sequence of data cells, each containing an individual data packet, embedded across a continuous time interval of an audio presentation. However, the present document does not set forth any requirement regarding the duration of the time interval that is embedded nor the alignment of its starting or ending boundaries with a logical boundary of the audio presentation or of the broadcast service (e.g. show segment, advertisement, etc.).

5. SPECIFICATION

5.1 Physical Layer

The physical layer properties of the audio watermark are described using continuous-time nomenclature, which has an unambiguous correspondence to discrete-time sampled system nomenclature given the audio sampling rate.

5.1.1 Marked Audio Signal

We define a marked audio signal as $s_i(t)$ where t is the continuous time variable and i is the audio signal index.

The audio watermark shall be present in the marking frequency band of the audio signal which is specified below in Table 5.2.

Let the subband of the marked audio signal $s_i(t)$ which lies in the marking frequency band be denoted $s_i'(t)$. We define the autocorrelation difference function, $R_d(i,t,\tau)$ as

$$R_d(i,t,\tau) = \int_{t}^{t+T/2} s_i'(u)s_i'(u-\tau)du - \int_{t+T/2}^{t+T} s_i'(u)s_i'(u-\tau)du$$

where t is the time instant at which the calculation is performed, τ is the autocorrelation delay, and T is the symbol interval.

Two audio watermark signaling modes are specified, “standard signaling” and “inverse signaling.” For each signaling mode, the value of the autocorrelation difference function at time $t_n = nT$ shall determine the symbol value embedded in the nth symbol interval as shown below in Table 5.1. Unless otherwise specified, standard signaling mode shall be employed in marked audio.

Table 5.1 Mapping of Autocorrelation Difference Value to Symbol and Bit Values

Value of $R_d(i,t,\tau)$	Standard Signaling		Inverse Signaling	
	Symbol Value	Bit Value	Symbol Value	Bit Value
Greater than or equal to zero	+1	1	-1	0
Less than zero	-1	0	+1	1

The parameters of $R_d(i,t,\tau)$ shall be as specified in Table 5.2 below.

Table 5.2 Parameter Specifications

Parameter Description	Parameter Value
Autocorrelation Delay, τ	3 milliseconds
Symbol interval, T	1/106 seconds
Marking Frequency Band	2.5 kHz - 5 kHz

The magnitude of $R_d(i,t,\tau)$ employed for a symbol dictates the content-dependent tradeoff between imperceptibility of the audio watermark to the audience and the robustness of the audio watermark to distortion introduced by subsequent audio processing. This value is not normatively specified.

As informative guidance to implementers, acceptable results for typical use cases of the VP1 audio watermark have been demonstrated by an implementation of the technology that employed an average recommended strength for symbols of:

$$\operatorname*{avg}_{n} \sigma_{i,n} = 0.3,$$

where $\sigma_{i,n}$ denotes the strength of symbol n in audio signal i and is defined as:

$$\sigma_{i,n} = \left| \frac{2R_d(i,t_n,\tau)}{E_s(i,t_n-\tau)+E_s(i,t_n)} \right|.$$

where $E_s(i,t)$ is the energy in the marking frequency band over the symbol interval starting at time t, calculated according to:

$$E_s(i,t) = \int_t^{t+T} s_i'(u)^2\,du$$

Symbol errors may be present in marked audio signals conforming with this specification as necessary to achieve a desired level of imperceptibility in marked audio presentations.

Marked audio signals shall be present only in a marked audio presentation, as specified in Section 5.1.2.

5.1.2 Marked Audio Presentation

All audio signals of a marked audio presentation shall be synchronously embedded such that the $R_d(i,t,\tau)$ for each of the audio signals indicates the same symbol in the same signaling mode (i.e., standard or inverse) for the same $t = t_n$.

For proper detection, all audio signals that are intended to be decoded and rendered simultaneously must contain identical watermarks. If any of the audio signals in the audio presentation are marked, then all of the audio signals of that audio presentation shall contain identical watermarks. If there are shared audio signals between two audio presentations, and any of the audio signals of the two audio presentations are marked, then all of the audio signals of both audio presentations shall contain identical watermarks.

5.2 Data Link Layer

5.2.1 Cell

A specific segment of the series of contiguous bits derived from the series of symbols in the audio signal at increasing values of $t = t_n$ shall form a cell of 159 bits which are ordered left to right, corresponding to the time-ordering of the symbols (first to last) across a 1.5 second interval of the audio signal.

The cell() structure shall have the syntax shown below in Table 5.3. It is divided into two regions: header and packet.

Table 5.3 Syntax cell() Structure

Syntax	No. of Bits	Format
cell() {		
header	32	bslbf
packet	127	bslbf
}		

header – A 32-bit header sequence as specified in Section 5.2.1.1, below.

packet – A 127-bit sequence conveying a packet of data.

5.2.1.1 Header

The first 32 bits of a cell is the header sequence, which shall have the value as shown below in Table 5.4.

Table 5.4 Header Sequence

	Header
Binary	10101110000010101011100111100100

5.2.1.2 Packet

The last 127 bits of a cell is the packet sequence, which shall convey a 127-bit packet of data.

A specification of the contents of a packet with application of the VP1 audio watermark to the recovery of ATSC 3.0 service signaling by redistribution receivers is provided in A/336 [3].

– End of Document –